1P-01

機械学習によるRGB画像からの距離画像の生成

佐藤颯人† 田村仁† 檜山 正樹‡ 入江 俊‡ 仲田 仁‡
日本工業大学 創造システム学科† 日本工業大学 機械システム工学専攻‡

1. はじめに

近年では距離画像センサが安価に普及し、モーションキャプチャやARの隠れ処理などへ応用されている．例えばモーションキャプチャに応用されたものが，Microsoft社から発売されているKinect v2 for Xbox Oneなどがある．しかし個人用携帯機器への搭載は一部を除いて行われていない．スマートフォンなど広く普及した携帯機器では単眼のRGBカメラが搭載されていることが一般的である．そのため，一般的な携帯機器では距離画像が取得できず，モーションキャプチャやARの隠れ処理が行えない．

本研究では距離画像センサを搭載していない機器から距離画像を獲得するため，近年発達してきた機械学習手法，そのなかのディープラーニングを用いた画像合成手法を使用して距離画像を生成することを目的とする．画像合成手法には，GAN[1]と呼ばれる画像合成手法やPix2Pix[2]といった画像変換などがある．この手法を応用し，同一視点で同時撮影されたRGB画像と距離画像の対を学習データとした機械学習を行えば，学習されたニューラルネットを用いて任意のRGB画像に対して距離画像を合成することができる．

2. 研究目的

本研究では機械学習を用いてRGB画像から距離画像を生成することを目的とする．また、将来的にはスマートフォンでの実装を目指す．

また，生成されたディープラーニングモデルの評価を次のように行う．

(1) 生成画像の評価
生成された画像と教師データであるdepth画像のピクセル単位での標準誤差．

(2) 学習時間の評価
学習データを一定回数学習させたときの学習にかかった時間．

(3) リアルタイム性の評価
実際に画像生成を行った時のリアルタイム性(ディープラーニングではGPUを用いることにより計算時間が大幅に減少するが本研究ではこのシステムをスマートフォンへ搭載することなどを考えてCPUを用いた場合の計算時間も比較する)

Generation of Depth image from RGB image by machine learning
Hayato Sato†, Hitoshi Tamura†, Masaki Hiyama‡, Suguru Irie‡ Hitosi Nakata‡
Nippon Institute of Technology Innovative Systems Engineering†, Nippon Institute of Technology Mechanical Systems Engineering Major ‡

3. 実験方法の検討

本研究ではSceneNet RGB-D[3]というphoto画像(RGB画像) とinstance画像とdepth画像(距離画像)がセットになったデータセットを使用した．このデータセットは実際の空間ではなく物理シミュレーションソフトで作成された3D空間上の室内にランダムに家具を置き，それを撮影したものである．このデータセットの構成は，親フォルダ以下に0～999までのフォルダ，またその下にdepth，instance，photoというフォルダがあり，320×240の画像が各300枚ずつ，計90万枚の画像がフォルダ分けされて入っている．本研究では3種類の画像の内，depthとphoto画像を使用し，この2種類の画像をpix2pixで使用できる形に変換し，変換器を作成した．

4. 実装

本実験ではaffinelayerのGitHubのpix2pix-tensorflow[4]リポジトリを使用した．

このデータセットはそのままではpix2pixに使用できないので，以下のプログラムを作成し実験用のデータセットを作成した．

(1) 0～999のフォルダ下にあるphoto，depthフォルダの中の画像を読み込み，番号を振りなおした後新規に作成したcolor，depthフォルダにコピーする．

(2) color，depthそれぞれの画像を読み込み，320×240の画像の中央を基準に240×240に切り取り，その画像を256×256の画像に変換する．その後2つの画像を連結させて1つの画像にして，新規に作成したtrainフォルダに保存する．

このプログラムを用いて30万枚のデータセットを作成した．

本実験で使用するディープラーニングの画像生成モデルの入力モデルはRGB画像であり，出力は距離画像になる．

5. 実験結果

本実験は以下の環境で行った.

OS	Windows 10 Enterprise
CPU	Intel(R)Core(TM)i3-3220CPU@ 3.30GHz
メモリ	8GB
GPU	GeForceGTX1050 Ti

実験を行うにあたり，テストデータの 30 万枚の画像をすべて使うと学習終了までに膨大な時間がかかってしまうため，データセットの枚数を少なくして複数の変換器を作成した．その後教師データの距離画像と生成された距離画像の誤差を算出し，その結果からテストデータ全体での標準誤差を算出することによって変換器の正確性の目安とした．またテスト時には共通のデータセットを使用した.

数回の実験の結果から学習に使用する画像は時間も考慮し１万枚とした．また，誤差が減らない原因としてデータセットの画像に似たような画像が多くあり，過学習が起きているのではないかと考えられたため，データセットからランダムに一万枚の画像を取得するプログラムを作成し，再度実験を行った.

(1)生成画像の評価

実験結果は次の通りである.

表１　実験結果

学習回数	標準誤差	分散
100 回	0.169	2.854
300 回	0.158	2.499
400 回	0.128	1.646
500 回	0.159	2.524
600 回	0.127	1.610

上記の結果から学習回数を重ねるごとに誤差が小さくなっていく傾向がみられた.

次に，学習回数 600 回の変換器を使用し，距離画像を生成した．(図 1)また，距離画像は明度を調整した.

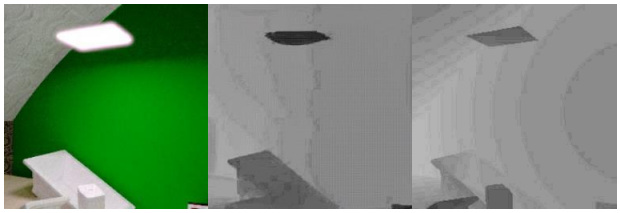

図 1　(左)変換器に入力した画像，(中)変換器から出力された画像，(右)本来の距離画像

実験の結果元の距離画像に近い画像が生成でき，ある程度の距離を測ることができた．このことから簡単な隠れ処理などの用途には使用できると考えられる．しかし近すぎるとうまく変換できない，何もないところが近いと表示されるなどうまく生成できない部分があった.

(2)学習時間の評価

本実験での学習時間は，学習用画像を 1 万枚用意したとき，学習回数 100 回で約 80 時間である.

(3)リアルタイム性の評価

今回使用したプログラムの画像一枚当たりの計算時間は次の通りである.

表２

使用演算装置	1 枚当たりの計算時間
GPU	約 0.067 秒
CPU	約 0.468 秒

実験の結果，スマートフォンへの搭載を考えたとき現状の計算速度では約 2fps しか得られないため，リアルタイム性に欠ける結果となった．しかし，将来スマートフォンに GPU が搭載されれば十分な速度を確保可能になると考えられる.

6. まとめ

機械学習を用いて RGB 画像から距離画像が生成できないかと考えた．pix2pix を用いた実験の結果，元の距離画像に近い画像が生成できた．しかし，数多くの問題が残っている.

この改善方法として，入力画像を小さくする，モデルのフィルター数を少なくするなどの方法が考えられる.

今後はこれらの問題点を改善し，より精度の高い距離画像の生成、及びスマートフォンでの実装を目指す.

参考文献

[1] Ian G.et al. "Generative adversarial nets Advances in neural information processing systems", pp. 2672–2680, 2014.
[2] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks". arXiv preprint arXiv:1611.07004 (2016).
[3] John McCormac and Ankur Handa and Stefan Leutenegger and Andrew J. Davison. 2016. SceneNet RGB-D: 5M Photorealistic Images of Synthetic Indoor Trajectories with Ground Truth. (https://robotvault.bitbucket.io/scenenet-rgbd.html) (accessed 2018-1-10)
[4](https://github.com/affinelayer/pix2pix-tensorflow)(accessed 2018-1-10)